

Contribution to the study and application of model driven and data driven engineering

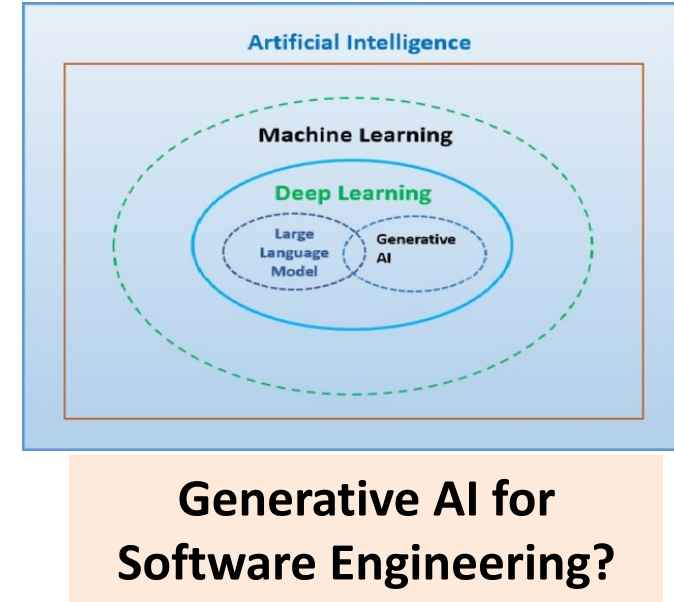
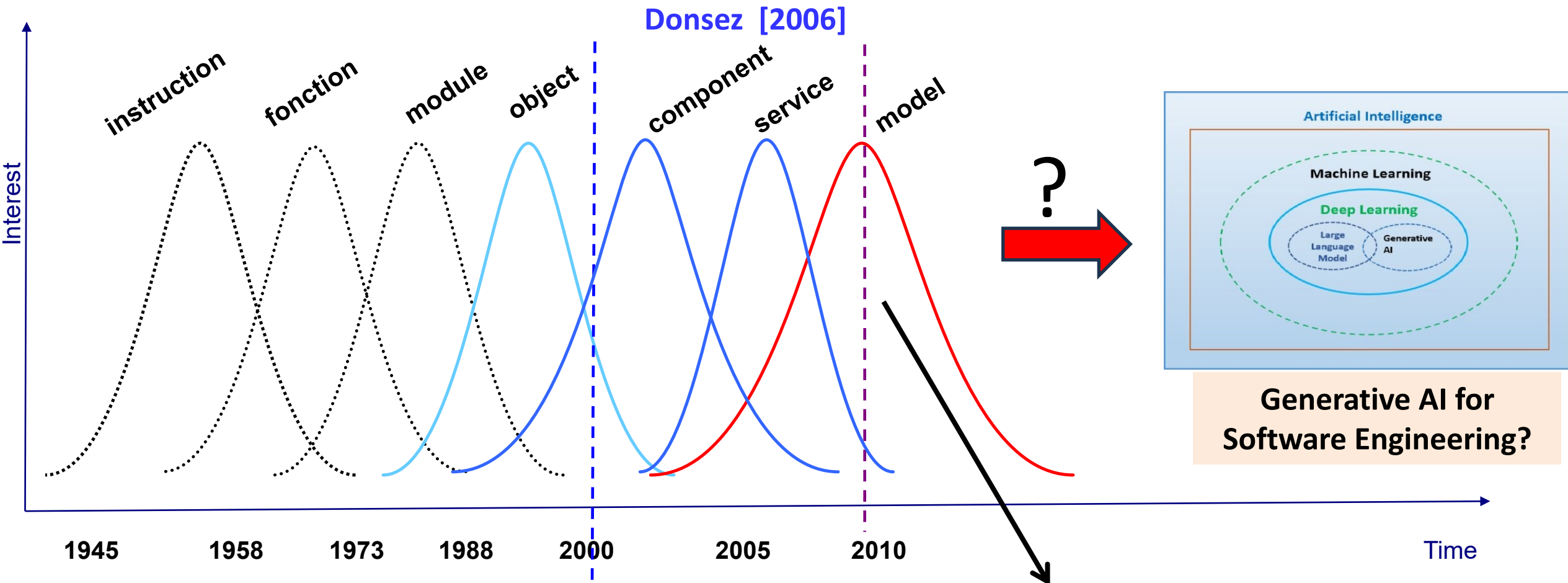
Slimane HAMMOUDI (HdR, 2010)
ESEO, electronics and computer engineering school
Angers, France

Application domains:

- **Business process management** → modeling, recommendation and reuse
- **Mobility and Smart Cities** → profiles, prediction and services recommendation
- **Big Data, NoSql Systems and Data Warehouse/Lake** → data quality, migration

Model Driven Engineering

■ Raccoon Waves [ACM, SIGSOFT, 1997]

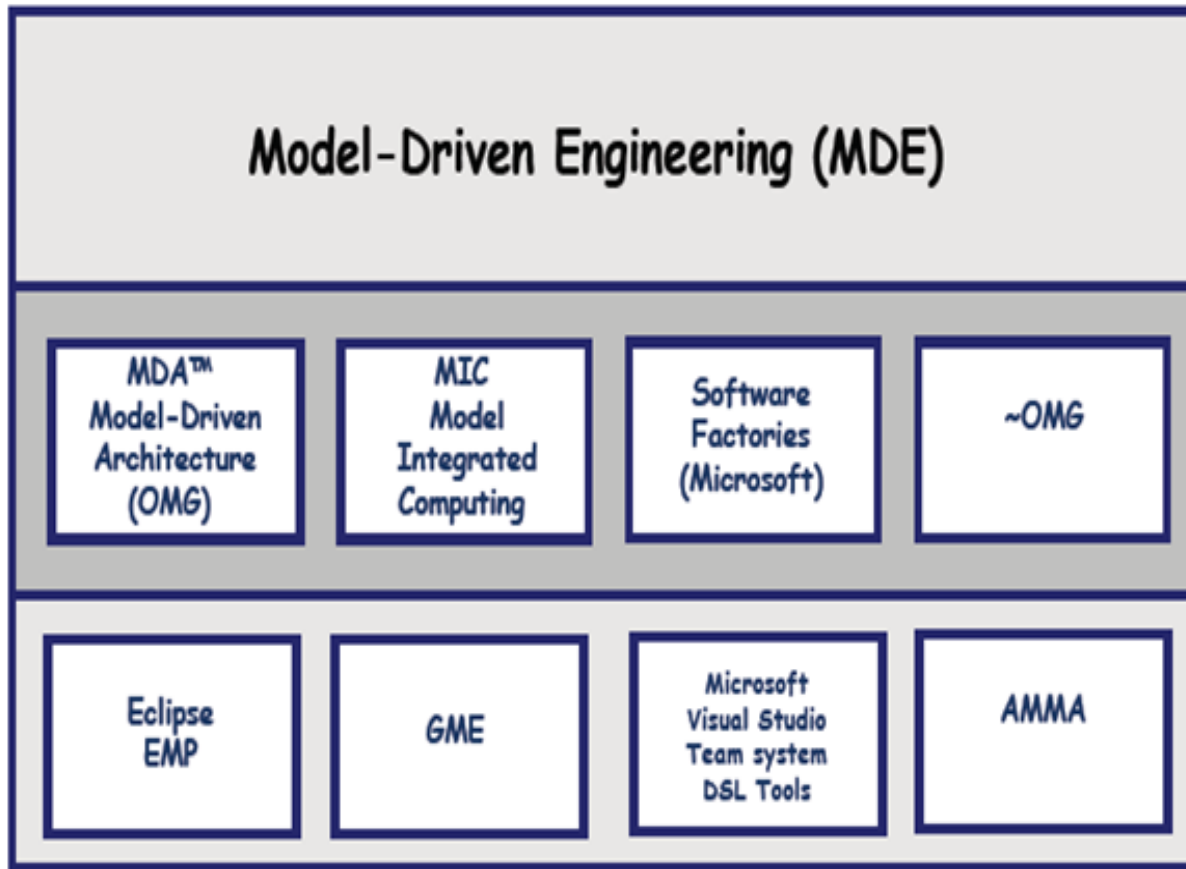


Fifty Years of Progress in Software Engineering

- ✓ Model Driven Engineering (MDE)
- ✓ Model Driven Architecture (MDA) – OMG [2000]

Model Driven Engineering

"A model is a simplification of a system built with an intended goal in mind. The model should be able to answer questions in place of the actual system " [Bézivin, 01].



Trend, approaches and platforms

main concepts:

- System
- Model
- Metamodel
- Transformation

main models (MDA):

- CIM
- PIM
- PSM
- Code

The main challenges:

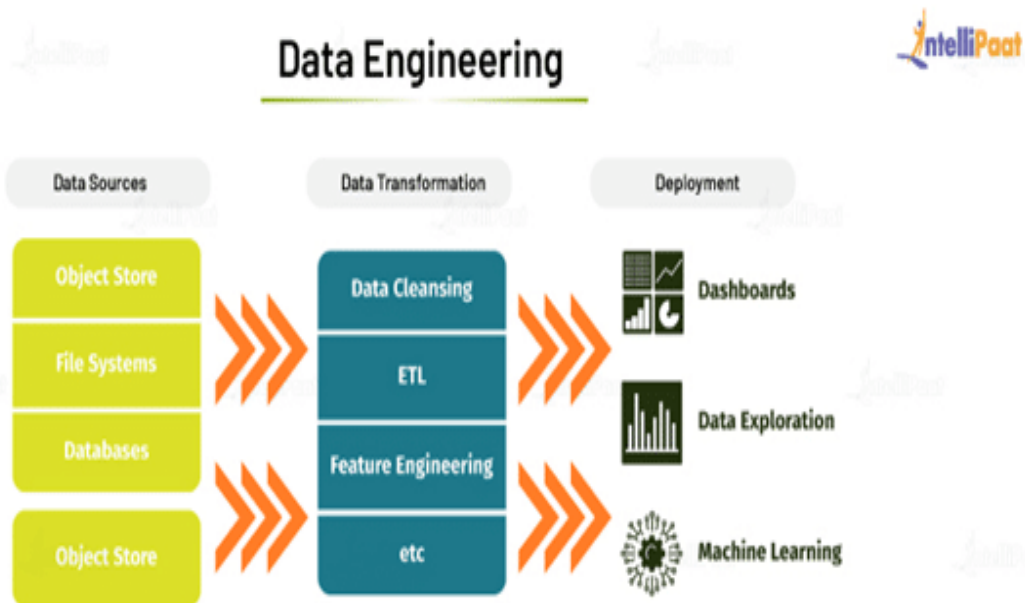
Dealing with the complexity of software development by raising the abstraction level and introducing more automation.

- Model transformation
- Run-time models and model quality
- Domain-specific modeling
- Model verification and validation

IS perspective

“Data engineering is the discipline of making suitable data accessible and available to different types of data consumers (including data scientists, business analysts, data analysts and others). » **@Gartner, 2020**

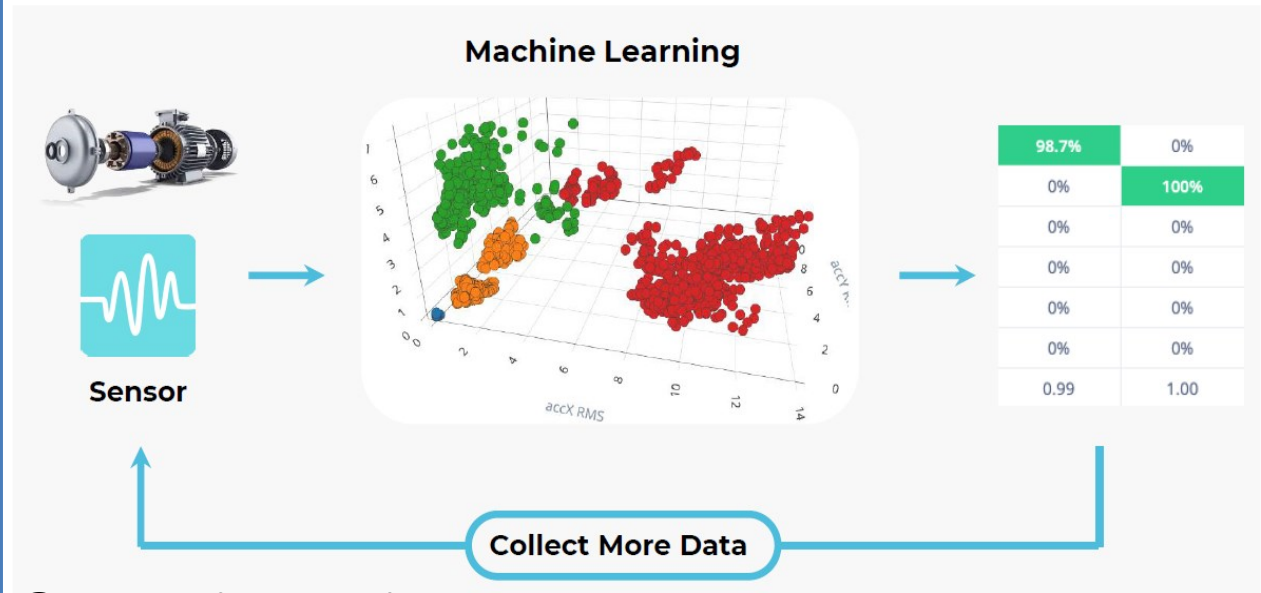
The foundation of data engineering from the information system perspective lies in the creation of **data pipelines**.



AI perspective

The data-driven way from AI perspective focusses on building a system that can identify what is the right answer based on having “seen” a large number of examples of question / answer pairs and “training” it to get to the right answer. **@ronald_istos**

The necessary components for this approach are an **appropriately large dataset** that, is also correctly labelled.



3 PhD

- ❑ MDA/MDE for Web services Platform
- ❑ MDA/MDE for Context_aware services
- ❑ Mapping versus Transformation in MDA/MDE

	IDM (MDA)	Services Web	Sensibilité au contexte
Modélisation & Métamodélisation	UML (généraliste) versus EDOC (DSL : langage spécifique au domaine)	Métamodèles pour les plateformes J2EE et dotNET Métamodèles UML versus EDOC	<i>modèle métier versus modèle de contexte</i> Métamodèle de contexte CPIM et CPSM
Transformation	<i>mapping versus transformation</i> <i>matching versus mapping</i> Processus semi-automatique	<i>application du principe mapping versus transformation</i> Transformation UML vers BPEL	Transformation paramétrée Composition de modèles
Méthodologie de développement	Approche MDA avec le principe (<i>mapping versus transformation</i>) Processus semi-automatique	MDA pour la plateforme services Web (<i>mapping versus transformation</i>)	Méthodologie COMODE (MDA pour applications sensibles au contexte)

Co-encadrement, Thèse de doctorat 2016, Université de Sfax, Tunisie, Prof. FEKI

"Techniques de *matching* entre métamodèles dans l'approche dirigée par les modèles
MDA : étude, évaluation et proposition."
(L. Lafi)

Co-encadrement, Thèse de doctorat 2017, Université de Tlemcen, Algérie, Prof. Chikh

Approche dirigée par les modèles pour le développement des applications *sensibles au*
contexte à base de services
(B. Boudaa)

Collaboration, Thèse de doctorat 2019, ESEO, UAngers, Mme Amghar (MC, HdR)

Recommandation contextuelle de services : application à la recommandation
d'évènements culturels dans la *ville intelligente*
(N. Gutowski)

Background:

Recommendation of relevant services to mobile user in a city is one of the main challenges in smart cities applications. In the last decade, mobility prediction has played a crucial role in urban planning, traffic forecasting, advertising, and service recommendation. In this PhD project, we focus on mobility prediction for service recommendation.

Approach:

To achieve better mobility prediction result, three main challenges must be taken into account: trajectory modelling, data processing and prediction process. Evaluation of our contributions is performed on a real dataset. First results showed the positive impact of our trajectory model on our mobility prediction process.

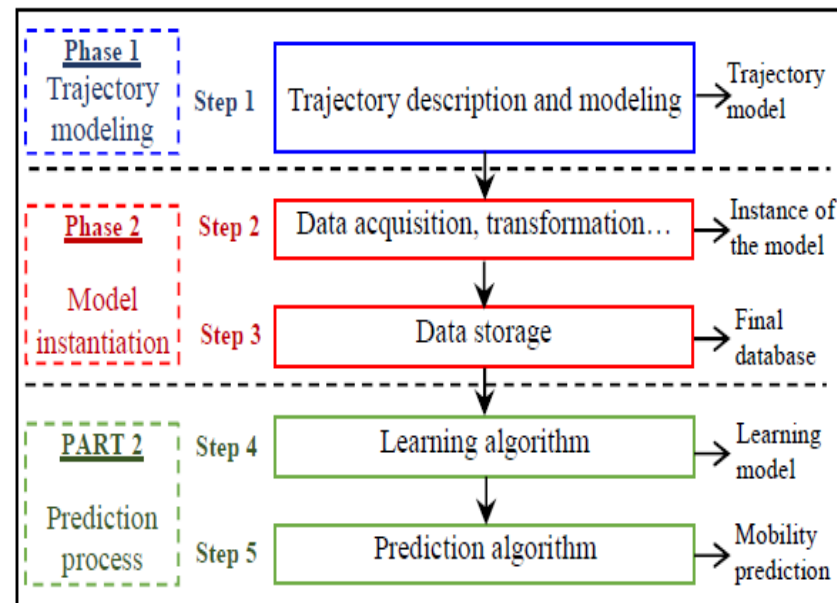


Fig. 1. Main steps of our mobility prediction approach



Background:

Modeling business process (BP) is an important issue in Business Process Management (BPM). However, modeling BP from scratch is fallible, complex, time-consuming and error prone task. BP model reuse has been recommended to reduce modeling time, decrease costs, and improve the quality of designed models. BP model repositories have been proposed to store collections of BP models for future reuse.

Approach:

A model driven engineering (MDE) approach and a recommendation process for BP model reuse is proposed. The recommendation process is based on the user social profile and implemented as a transformation process in model driven engineering. In our experiments, the **LinkedIn social network** is used to extract the users' business interests. These user business interests are then used to recommend the appropriate BP models that could fit to the user.

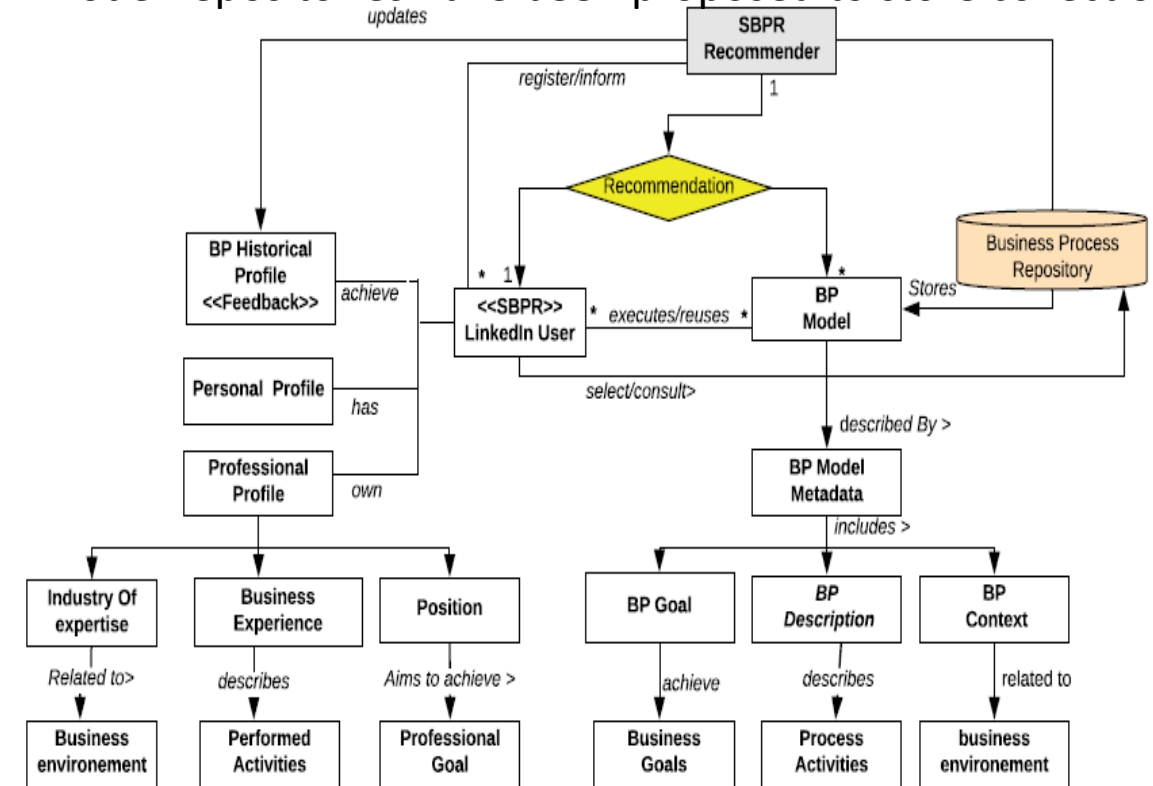


Figure 1: SBPR Framework

*Social Business Process Recommender

Background: The data warehousing process requires an architectural revolution to settle Big Data challenges and address new data sources. In this respect, the pipeline modeling community for the acquisition, storage, and processing of data for analysis purposes is enacting a wide range of technological solutions that present significant challenges and difficulties. More specifically, the choice of the most appropriate tool for the user's specific business needs and the interoperability between the different tools became a primary challenges.

Approach: An interactive framework based on ML techniques to assist experts in the process of modeling a customized pipeline for data warehousing is proposed.

- 1- We elaborated an analysis of the experts' requirements and the characteristics of the data to be processed.
 - 2- We proposed the most appropriate architecture to their requirements from a multitude of specific architectures instantiated from a generic one.
 - 3- We used several ML methods to predict the most suitable tool in each architecture' phase and task.
- Finally, our framework is validated through two real world use cases and the users feedback.

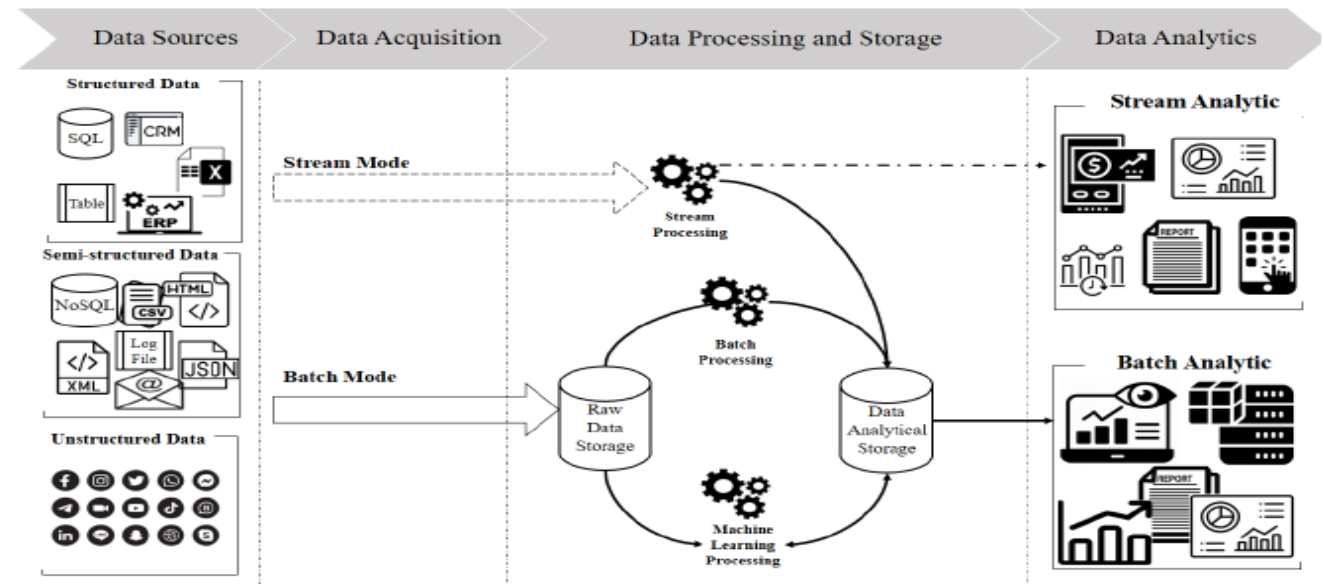


Fig. 1: The generic architecture: Tool Independent Architecture level.

2016 – 2023 : TRIMANE – IRIT, Toulouse (Gilles Zurfluh, Fatma Abdelhedi)

- ❑ Modélisation de bases de données NoSQL : Processus d'extraction des schémas logique et conceptuel (**Hela Rajhi, Décembre 2023**) → *Allel Hadjali, Rapporteur*
- ❑ Processus d'ingestion de données hétérogènes et d'assistance au requêtage pour un lac de données médical (**Rym Jemmali, Décembre 2023**) → *Slimane Hammoudi, Rapporteur*
- ❑ Extraction des modèles d'une base de données NoSQL orientée-documents basée sur une approche dirigée par les modèles (**Rabah Tighilt Ferhat, Novembre 2021**) → *Slimane Hammoudi, Rapporteur*

2024 – 2028 : TRIMANE – ESEO – LIAS (Allel Hadjali, Slimane Hammoudi)

- ❑ Deux axes de recherche:
 - Modélisation et Rétro ingénierie des bases de données NOSQL
 - Conception centrée qualité de Lac de données

- ❑ Thèse Cifre (**Mai 2024**):

Modèle générique de métadonnées centré qualité pour les Data Lake : Application aux données de santé

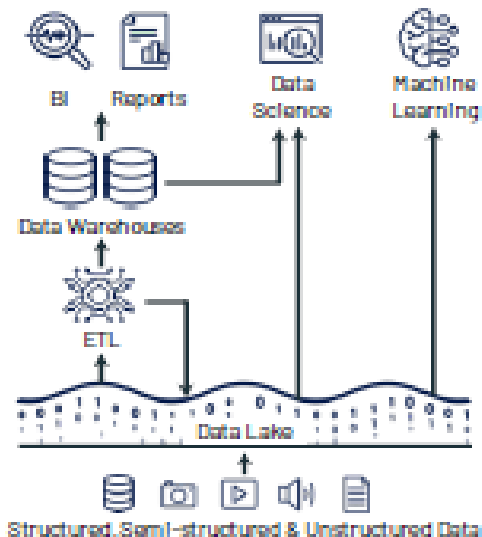
Background

The study of data lake metadata models/systems is currently an active research topic and many proposals have been made over the last 4 years.

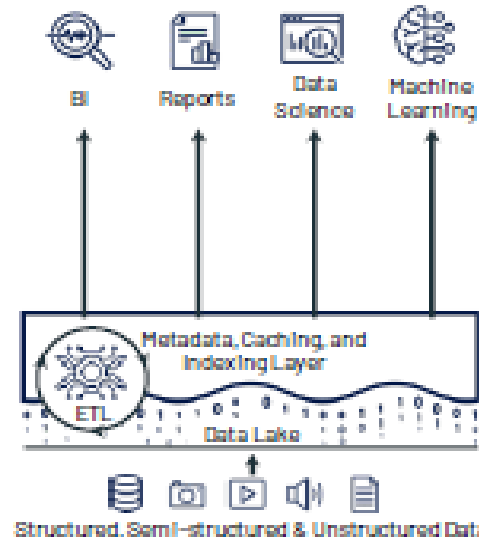
- The proposed models are not generic enough to handle different types of data lakes (SQL, NoSQL, XML, JSON,...)
- They didn't include criteria to ensure the quality and integrity of data in a data lake, such as entity resolution and disambiguation.



(a) First-generation platforms.



(b) Current two-tier architectures.



(c) Lakehouse platforms.

Evolution of data platform architectures to today's two-tier model (a-b) and the new Lakehouse model (c) [Armbrust; 2021]

Research issues:

- How to identify and characterize all the dimensions (structural, semantic, terminological, extensional) to address **qualities issues such as entity resolution and disambiguation** in a Data Lake?
- How to define a **generic quality-centered metadata model** to handle different types of data lakes and to ensure quality and integrity of data in a data lake?
- How to define a **functional architecture for the implementation of a Data Lake** taking into account the proposed metadata model in order to ensure quality criteria?

Case study:

- The health insurance fund has developed a Digital Health Space (named ENS¹). The ENS constitutes a real Data Lake because of (1) the diversity of data types, media and formats (2) the volumes stored which can reach several terabytes and (3) the raw nature of the data

¹Espace numérique de santé

Related Work :

Modèles \ Caractéristiques	ES	PZ	VD	SU	CG	LS	PM	GM	Total
GEMMS [Quix et al., 2016]	✓				✓		✓	✓	4/8
Ground [Hellerstein et al., 2017]	✓		✓	✓	✓		✓		5/8
[Diamantini et al., 2018]	✓	✓				✓		✓	4/8
[Ravat and Zhao, 2019b]	✓	✓	✓	✓	✓	✓	✓		7/8
HANDLE [Eichler et al., 2020]	✓	✓		✓	✓	✓	✓	✓	7/8
MEDAL [Scholly, 2022a]	✓	✓	✓	✓	✓	✓	✓		
goldMEDAL [Scholly, 2022b]	✓	✓	✓	✓	✓	✓	✓	✓	
Total	7/7	5/7	4/7	5/7	6/7	5/7	6/7	4/7	

Features supported by data lake metadata models [Scholly, 2023]

Data Lakes: A Survey of Functions and Systems

IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING,
VOL. 35, NO. 12, DECEMBER 2023

Rihan Hai , Christos Koutras , Christoph Quix , and Matthias Jarke

A Vision for Data Alignment and Integration in Data Lakes

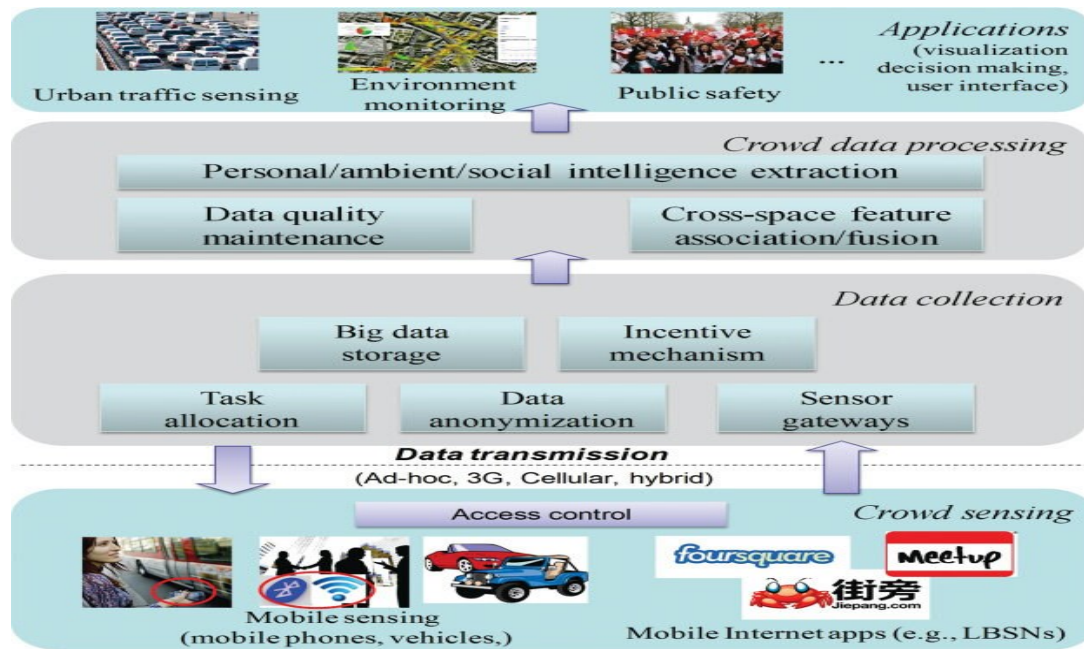
Keynote, IEEE Big Data, Osaka, 2022

Renée J. Miller, Professor, Northeastern University, USA

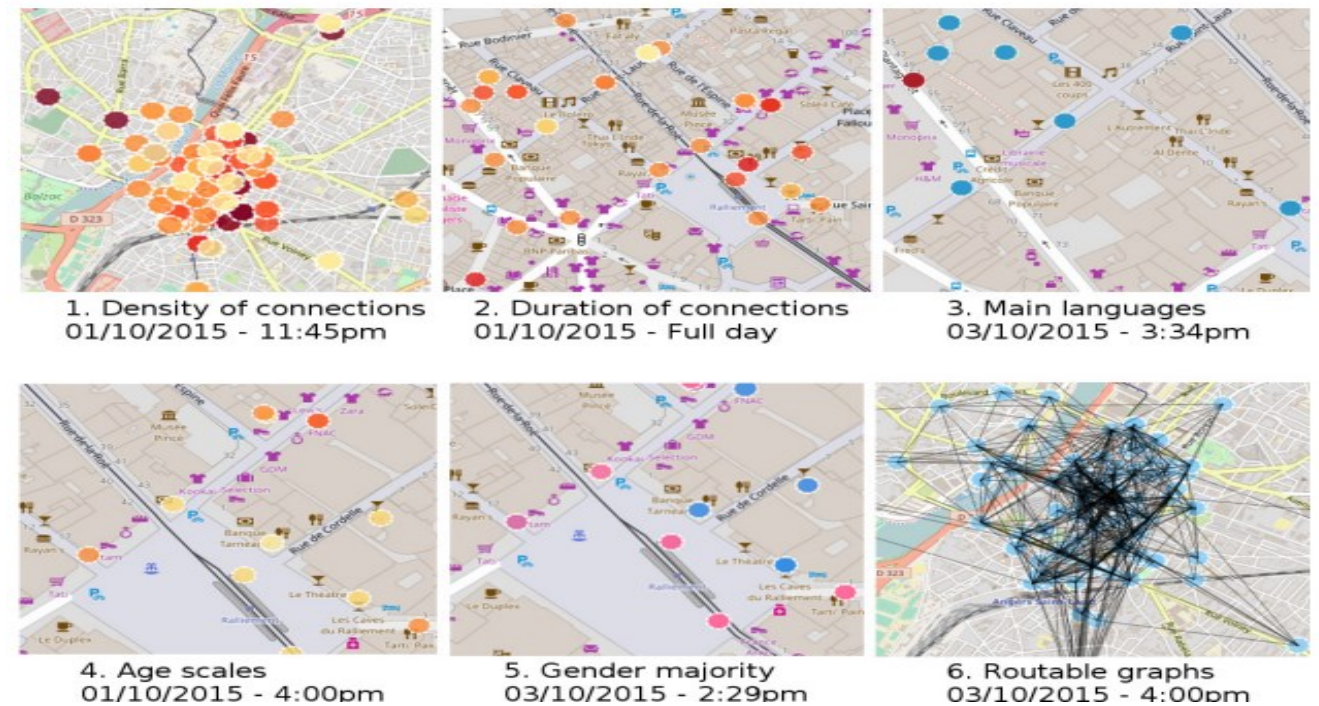
Background:

This research project addresses two important challenges for the design of a platform dedicated to the analysis of citizens mobility in a smart city:

- Predicting individual mobility patterns from mobile network data (WIFI, GSM)
- Prevention of pollution in a city based on the analysis and prediction of citizen mobility.



Mobile Crowd Sensing and Computing
[Guo & al, 2015]



WifiLib connections in Angers [Gutowski & al, 2019]

Research issues:

- How to define a **context model** which covers data of the city and its environment in order to carry out relevant individual mobility patterns predictions?
- How to define a **privacy aware trajectory model** for mobile user displacements in an Urban environment.
- Which **prediction techniques** should we use to find the possible mobile user trajectories?
- Which **methodology based on the analysis/prediction** of citizens mobility to design a platform to ensure pollution prevention in the different areas of the smart city.
- Which **“Mobility-Centric” and Trustworthy Internet Architecture** to design a platform to ensure pollution prevention and mobility prediction in smart cities.

Case study:

- **Dataset** : *Wifilib* Angers. It contains mobility data and some profile data of about 80K citizens, connected to the network, during the year 2017. 219 access points, 79153 users and 15139162 connections to the network
- **Free Wi-Fi in Algiers, 3,616** *Free Wi-Fi hotspot*

Related/Previous Work :

« Mobility Prediction, Context and Trajectory modeling »

- 1. Hocine Boukhedouma, Abdelkrim Meziane, Slimane Hammoudi, Amel Benna and Hadjali Allel.**
Towards a Context-based Mobility Prediction in Smart Cities: First Experimentations. *The 11th IEEE International Conference on Smart City and Informatization. England, UK, 1-3 November 2023*
- 2. Hocine Boukhedouma, Abdelkrim Meziane, Slimane Hammoudi and Amel Benna**
A Grid-based and a Context-oriented Trajectory Modeling for Mobility Prediction in Smart Cities, *At the 8th International Conference on Smart City Applications, Held from 4 to 5 October 2023 at the ESTP-Paris France.*
- 3. Hocine Boukhedouma, Abdelkrim Meziane, Slimane Hammoudi and Amel Benna**
On the Challenges of Mobility Prediction in Smart Cities. September 2020, 5th International Conference on Smart Data and Smart Cities, Nice France

« Mobility analysis, Air Pollution Monitoring »

Mohammad ABOUD

Leveraging Machine Learning for Multi-Source Data Enrichment and Analytics in Environmental Monitoring and Crowd Sensing
Thèse de doctorat de l'Université Paris-Saclay, Novembre 2023.

Wei Ying Yi, Kin Ming Lo, Terrence S. T. Mak, Kwong-Sak Leung, Yee Leung, Helen Mei-Ling Meng:

A Survey of Wireless Sensor Network Based Air Pollution Monitoring Systems. [Sensors 15\(12\)](#): 31392-31427 (2015)

National:

TRIMANE: Abdelhedi Fatma (Directrice de recherche, CBI2)

LIAS (Poitiers): Prof. Allel HADJALI

LISTIC (Chambery): Prof. Sébastien Monnet (Thèse Asma Dhaouadi)

International:

CERIST (Algeria): Prof. Meziane Abdelkrim

RIADI (Tunisia): Prof. Gammoudi Mohammed

IPS, Setubal (Portugal): Prof. Filipe Joaquim

FEMCS, Twente (Netherlands): Prof. Ferreira Pires Luis

UFMA (Brazil): Prof. Lopes Denivaldo

iDEA Lab (Italy): Prof. Cuzzocrea Alfredo



References 2023-2024

1. Asma Dhaouadi, Khadija Bousselmi, Mohamed Mohsen Gammoudi, Sébastien Monnet, Slimane Hammoudi
Machine Learning Based Decision Support Framework For Big Data Pipeline Modeling And Design
submitted to Jordanian Journal of Computers and Information Technology, March 2024. IF 2.06
2. Hadjer Khider, Slimane Hammoudi, Abdelkrim Meziane
Towards a model-driven engineering approach and a recommendation process for BP model reuse.
International Journal of Computer Languages Submitted, March 2024, Elsevier, IF 2.2
3. A. Dhaouadi, W.Paccoud, K.Bousselmi, S.Monnet, M.Gammoudi, S.Hammoudi,
Big Data Tools: Interoperability Study and Performance Testing.
IEEE Big Data workshop on Methods to Improve Big Data Science Projects (MIDP-2023) Sorrento, Italy, December 2023.
4. Ichou Sabrina, Hammoudi Slimane, Benna Amel, Alfredo Cuzzocrea, and Meziane Abdelkrim
Towards Big Data Analytics over Mobile User Data using Machine Learning.
IEEE BigData 2023 special session MLBD 2023. Sorrento, Italy, December 2023.
5. Hocine Boukhedouma, Abdelkrim Meziane, Slimane Hammoudi, Amel Benna and Hadjali Allel.
Towards a Context-based Mobility Prediction in Smart Cities: First Experimentations.
The 11th IEEE International Conference on Smart City and Informatization. England, UK, 1-3 November 2023
6. Hocine Boukhedouma, Abdelkrim Meziane, Slimane Hammoudi and Amel Benna
A Grid-based and a Context-oriented Trajectory Modeling for Mobility Prediction in Smart Cities,
At the 8th International Conference on Smart City Applications, Held from 4 to 5 October 2023 at the ESTP-Paris France.

